THE ANNALS OF "DUNAREA DE JOS" UNIVERSITY OF GALATI
FASCICLE III, 2000 ISSN 1221-454X
ELECTROTECHNICS, ELECTRONICS, AUTOMATIC CONTROL, INFORMATICS

GIS COMPONENTS: AN OVERVIEW OF THE SOFTWARE FORMAT APPROACHES

Daniela NEAGU[1], Ciprian-Daniel NEAGU[2]

[1]Department of Fishing and Aquaculture, "Dunarea de Jos" University of Galati
[2]Department of Computer Science & Eng., "Dunarea de Jos" University of Galati

Abstract: The paper proposes a software/ multimedia point of view of the resources and file formats used in GIS applications. An overview of the most known software approaches for GIS data management and components is proposed in order to familiarize and focus to the one of the most dynamic multimedia tools. A syntesis of the characteristics of specific file formats for geographic knowledge representation is made.

Keywords: GIS software, file format, GIS databases, geographic data representation.

1. INTRODUCTION

A geographic information system (GIS) is a computer-based tool for mapping and analyzing things that exist and events that happen on earth. GIS technology integrates common database operations such as query and statistical analysis with the unique visualization and geographic analysis benefits offered by maps. These abilities distinguish GIS from other information systems and make it valuable to a wide range of public and private enterprises for explaining events, predicting outcomes, and planning strategies. Whether siting a new business, finding the best soil for growing bananas, or figuring out the best route for an emergency vehicle, local problems also have a geographical component. GIS give the power to create maps, integrate information, visualize scenarios, solve complicated problems, present powerful ideas, and develop effective solutions like never before. GIS is a tool used by individuals and organizations, schools, governments, and businesses seeking innovative ways to solve their problems.

Mapmaking and geographic analysis are not new, but a GIS performs these tasks better and faster than do the old manual methods. Before GIS technology, only a few people had the skills necessary to use geographic information to help with decision making and problem solving. Today, professionals in every field are increasingly aware of the advantages of thinking and working geographically.

1.1. What Can GIS Do

As a consequence of actual improvements in applied software in geographical information, GIS perform geographic queries and analysis, improve organizational integration, make better decisions, improve the process of making maps (DMA 1991).

GIS Perform Geographic Queries and Analysis

The ability of GISs to search databases and perform geographic queries have helped reduce costs by: streamlining customer service, reducing land acquisition costs through better analysis, reducing fleet maintenance costs through better logistics, analyzing data quickly (example: a realtor could use a GIS to find all houses within a certain area that have tiled roofs and five bedrooms, then list their characteristics, DMA 1977).

Improve Organizational Integration

Many organizations that have implemented a GIS have found that one of its main benefits is improved management of their own organization and resources. Because GISs have the ability to link data sets together by geography, they facilitate interdepartmental information sharing and communication. By creating a shared database, one department can benefit from the work of another - data can be collected once and used many times.

This paper was recomended for publication by S. Bumbaru

Make Better Decisions

The old adage "better information leads to better decisions" is as true for GIS as it is for other information systems. A GIS, however, is not an automated decision making system but a tool to query, analyze, and map data in support of the decision making process. GIS technology has been used to assist in tasks such as presenting information at planning inquiries, helping resolve territorial disputes, and siting pylons in such a way as to minimize visual intrusion (Neuner 2000).

GIS can be used to help reach a decision about the location of a new housing development that has minimal environmental impact, is located in a low-risk area, and is close to a population center. The information can be presented succinctly and clearly in the form of a map and accompanying report, allowing decision makers to focus on the real issues rather than trying to understand the data. Because GIS products can be produced quickly, multiple scenarios can be evaluated efficiently and effectively.

Making Maps

Maps have a special place in GIS. The process of making maps with GIS is much more flexible than are traditional manual or automated cartography approaches. It begins with database creation. Existing paper maps can be digitized and computer-compatible information can be translated into the GIS. The GIS-based cartographic database can be both continuous and scale free. This allows the creation of map products which are centered on any location, at any scale, and showing selected information symbolized effectively to highlight specific characteristics. The characteristics of atlases and map series can be encoded in computer programs and compared with the database at final production time. Digital products for use in other GISs can also be derived by simply copying data from the database. In a large organization, topographic databases can be used as reference frameworks by other departments.

1.2. How GIS works?

A GIS stores information about the world as a collection of thematic layers that can be linked together by geography. This extremely powerful and versatile concept has proven invaluable for solving many real-world problems from tracking delivery vehicles, to record details of planning applications, to model global atmospheric circulation.

Geographic References

Geographic information contains either an explicit geographic reference, such as a latitude and longitude or national grid coordinate; or an implicit reference such as an address, postal code, census tract name, forest stand identifier, or road name. An automated process called geo-coding is used to create explicit geographic references (multiple locations) from implicit references (descriptions, addresses). These geographic references allow an user to locate features such as a business or forest stand, and events, such as an earthquake, on the earth's surface for analysis.

Vector and Raster Models

Geographic information systems work with two fundamentally different types of geographic models - the "vector" model and the "raster" model. In the vector model, information about points, lines, and polygons is encoded and stored as a collection of *x,y* coordinates. The location of a point feature, such as a bore hole, can be described by a single *x,y* coordinate. Linear features, such as roads and rivers, can be stored as a collection of point coordinates. Polygonal features (river catchments, sales areas) can be stored as a closed loop of coordinates.

The vector model is extremely useful for describing discrete features, but less useful to describe continuously varying features (soil type, accessibility costs for hospitals). The raster model has evolved to model such continuous features. A raster image comprises a collection of grid cells rather like a scanned map or picture. These software models for geographic data storing have advantages and disadvantages. Modern GISs are able to handle both models.

1.3. Components of a GIS

A working GIS integrates five key components: hardware, software, data, people, and methods. Hardware is the computer on which a GIS operates. Today, GIS software runs on a wide range of hardware types, from centralized computer servers to desktop computers used in stand-alone or networked configurations.

Software

GIS software provides the functions and tools needed to store, analyze, and display geographic information. Key software components are: tools for the input and manipulation of geographic information; a database management system (DBMS); tools that support geographic query, analysis, and visualization; a graphical user interface (GUI) for easy accessing.

Data

Possibly the most important component of a GIS is the data. Geographic data and related tabular data can be collected in-house or purchased from a commercial data provider. A GIS will integrate spatial data with other data resources and can even use a DBMS, used by most organizations to organize and maintain their data, to manage spatial data.

People

GIS technology is of limited value without the people who manage the system and develop plans for applying it to real-world problems. GIS users range from technical specialists who design and maintain the system to those who use it to help them perform their everyday work (Muehrcke 1992).

Methods

A successful GIS operates according to a well-designed plan and business rules, which are the models and operating practices unique to each organization (Snyder 1987).

1.4. Related Technologies

GISs are closely related to several other types of information systems, but it is the ability to manipulate and analyze geographic data that sets GIS technology apart. Although there are no hard and fast rules about how to classify information systems, the following discussion should help differentiate GIS from desktop mapping, computer-aided design (CAD), remote sensing, DBMS, and global positioning systems (GPS) technologies.

Desktop Mapping

A desktop mapping system uses the map metaphor to organize data and user interaction. The focus of such systems is the creation of maps: the map is the database. Most desktop mapping systems have more limited data management, spatial analysis, and customization capabilities. Desktop mapping systems operate on desktop computers such as PCs, Macintoshes, and smaller UNIX workstations.

CAD

CAD systems evolved to create designs and plans of buildings and infrastructure. This activity required that components of fixed characteristics be assembled to create the whole structure. These systems require few rules to specify how components can be assembled and very limited analytical capabilities. CAD systems have been extended to support maps but typically have limited utility for managing and analyzing large geographic databases.

Remote Sensing and GPS

Remote sensing is the art and science of making measurements of the earth using sensors such as cameras carried on airplanes, GPS receivers, or other devices. These sensors collect data as images and provide specialized capabilities for manipulating, analyzing, and visualizing those images. Lacking strong geographic data management and analytical operations, they cannot be called true GISs.

DBMS

Database management systems specialize in the storage and management of all types of data including geographic data. DBMSs are optimized to store and retrieve data and many GISs rely on them for this purpose. They do not have the analytic and visualization tools common to GIS.

Spatial Relationships and Technology

When viewing a map, the map-reader must interpret a variety of points, lines, and other symbols to identify spatial relationships among the geographic entities represented. For example, a map is used to find a route from one city to another, or to identify which county contains a feature of interest. The information required to perform these analyses is not explicit in the map; rather, the map-reader must interpret the required spatial relationships from mapped objects.

In a GIS database, the method by which spatial relationships are explicitly represented is termed topology. Topology is used to describe how linear objects connect, to define areas, and to identify the areas lying to either side of a linear object. Information about these spatial relationships is stored in a topological data structure and is essential to carry out most GIS functions.

2. GIS DATA FORMATS

2.1. Digital Map Formats

The term file format refers to the logical structure used to store information in a GIS application file. File formats are important in part because not every GIS software package supports all formats. If the user want to use a data set, but it isn't available in a format that the GIS supports, it is necessary to find a way to transform it, find another data set, or find another GIS.

Almost every GIS has its own internal file format. These formats are designed for optimal use inside the software and are often proprietary. They are not designed for use outside their native systems. Most systems also support transfer file formats. Transfer formats are designed to bring data in and out of the GIS software, so they are usually standardized and well documented.

If user data needs are simple, the main concern will be with the internal format GIS software supports. If the user have complex data needs, a wider range of transfer formats are required, especially if it needs to mix data from different sources. Transfer formats will be required to import some data sets into the software.

2.2. Vector Formats

Many GIS applications are based on vector technology, so vector formats are the most common. They are also the most complex because there are many ways to store coordinates, attributes, attribute linkages, database structures, and display information. Some of the most common formats are briefly described below and summarized in Table 1.

Arc Export

Arc Export is a transfer format, either ASCII or compressed into binary used to transfer files between different versions of ARC/INFO. It is undocumented and will work only with ESRI products.

AutoCAD" Drawing Files (DWG)

DWG is the internal, proprietary format used in AutoCAD® software, which is a computer-aided design/drafting (CAD) program. AutoCAD can convert any DWG file to a DXF file (described below) without loss of graphic information. As with DXF files, there are a number of ways to store attribute information in DWG files. The emerging standard is one that uses Extended Entity Data (EED) to link attributes, but many others are possible. The lack of one standard for linking attributes can cause problems when data is transferred between systems.

Autodesk's Data Interchange File (DXF) Format

DXF is probably the most widely used vector data transfer format, and a file in DXF format offers some very strong advantages. It contains very complete display information, and almost every graphics program can read it. There are several different ways to store attribute information in DXF and to link DXF entities to external attributes. Because there are no attribute standards, many programs that claim to read DXF files still do not import attribute information properly.

Digital Line Graphs (DLG)

DLG, a transfer format used by the US Geological Survey (USGS), depicts vector information portrayed on printed paper maps. It carries accurate coordinate information and sophisticated feature-classification information but no other attribute data. DLG does not include any display information. The DLG standard is significant because the USGS and other US government agencies have used it to publish large numbers of digital maps.

Hewlett-Packard Graphic Language (HPGL)

HPGL is a language that controls computer plotters; it contains display information but no geographic coordinates or attribute data. It is usually not appropriate for the storage or transfer of GIS data.

MapInfo" Data Transfer Files (MIF/MID)

MIF/MID is a transfer standard used by MapInfo, a desktop mapping system. It carries all three types of GIS information: geographic, attribute, and display. Attribute links are implicit in the file format.

MapInfo Map Files

MapInfo has its own internal binary format, known as a map file. It is undocumented and proprietary, so it cannot be used outside a MapInfo system.

MicroStation Design Files (DGN)

DGN is the internal format used by Bentley Systems Inc.'s MicroStation (CAD program). It may be used as a transfer standard because it is well documented and standardized. DGN files contain detailed display information. A common way to store attributes is to place them in an external database file and record links in the MSLINK field, a data item carried for each element in the DGN file.

Spatial Data Transfer System (SDTS)

SDTS, a new transfer format developed by the US government, was designed to handle all types of geographic data. SDTS can be either binary or ASCII but is generally binary. All geographic concepts can be encoded in SDTS, including coordinate information, complex attribute information, display information. This versatility causes a corresponding increase in complexity. To simplify things, standard subsets of SDTS have been adopted. Topological Vector Profile (TVP), for example, is used to store certain types of vector maps, and can be used for raster information. As a NSDI (US National Spatial Data Infrastructure) foundation, its importance will increase as more data becomes available.

Topologically Integrated Geographic Encoding and Referencing Files (TIGER)

TIGER is an ASCII transfer format used by the US Census Bureau to store the street maps constructed for the 1990 census. It contains complete geographic coordinates and is line, not polygon, based (although polygons can be constructed from its attribute information). The most important attributes include street name and address information. TIGER does not contain display information.

Vector Product Format (VPF)

VPF is a binary format used by the US Defense Mapping Agency. It is well documented and can be used as an internal and transfer format. It carries geographic and attribute information but no display data. VPF files are also referred to as VMAP products. The Digital Chart of the World (DCW) is published in this format.

Table 1. Software File-Format Chart

Application	.tiff	.tfw header	.dxf	.dem	.eoo	.shp	.jpg	igds/.dgn	.mif	.dlg	.sdts	dted	tiger
AutoCad V.13	2	2	1	3	1	1	1	1	1				
AutoCad V.14	1	2	1	3	1	1	1	1	1				
Cad Overlay GSX	1	1	1										
Caliper Corp.													
GIS Plus	1	2	1	1									
Maptitude	1	2	1	1	1	1			1	1			1
Trancad	1	2	1	1									
				3									
ENVI	1	1	3										
ERDAS	1	1	1	1									
ERMAPPER	1	1	1	1	1		1	1		1		1	
ESRI													
ARC/INFO	1	1	1	1	1	1	1	1	1	1	1		1
ArcView	1	1	1	2		1	1		1				2
Data Automation Kit				1	1	1			1				1
Geographix	1	1	1										
Genasys	1	1	1	1									
GRASS	1	1	1	1	1								
IDRISI	1	1	1	3									
Intergraph													
IRAS C	1	3	3	3									
Vista Map	1	3	3	3									
Landform Gold	1	1	3										
Microstation				3									
Descartes	2	2	3	3									
				2									
MapInfo	1	3	1		1	1		1	1				
Terramodel	1	1	1	1									
TOPO+	1	1	3	3									
Vertical Mapper	2	3	2	2									
1= Full compatibility													
2= Compatible only with third party software													
3= Not compatible at all													

2.3. Raster Formats

Raster files are used to store image information, such as aerial photographs or scanned paper maps. They are also used for data captured by satellite and other airborne imaging systems. Images from these systems are often referred to as remote-sensing data. Unlike other raster files, which express resolution in terms of cell size and dots per inch (dpi), resolution in remotely sensed images is expressed in meters, which indicates the size of the ground area covered by each cell. Some common raster formats are described below and summarized in Table 1.

Arc Digitized Raster Graphics (ADRG)

ADRG is a format used by the US military to store raster images of paper maps.

Band Interleaved by Line (BIL)

Band Interleaved by Pixel (BIP), and Band Sequential (BSQ). BIL, BIP, and BSQ are formats produced by remote-sensing systems. The primary difference among them is the technique used to store brightness values captured simultaneously in each of several colors or spectral bands.

Digital Elevation Model (DEM)

DEM is a raster format used by the USGS to record elevation information. Unlike other raster file formats, DEM cells do not represent color brightness values, but rather the elevations of points on the earth's surface.

PC Paintbrush Exchange (PCX)

PCX is a common raster format produced by most scanners and personal computer drawing programs.

Spatial Data Transfer Standard (SDTS)

As was indicated under vector formats above, SDTS is a general-purpose format designed to transfer geographic information. One SDTS variant is the raster profile, designed as a standard format for transferring raster data. However, this protocol has not as yet been finalized.

Tagged Image File Format (TIFF)

Like PCX, TIFF is a common raster format produced by PC drawing programs and scanners.

2.4. Types of Information in a Digital MapRaster

Any digital map is capable to store much more information than a paper map of the same area, but it's generally not clear at first glance what sort of information the map includes. For example, more information is usually available in a digital map than what it is seen on-screen. Evaluating a given data set by simply looking at the screen can be difficult: what part of the image is contained in the data and what part is created by the GIS program's interpretation of the data?

Three general types of information can be included in digital maps: *geographic information* (which provides the position and shapes of specific geographic features); *attribute information* (which provides additional non-graphic information about each feature); *display information* (which describes how the features will appear on the screen).

Some digital maps do not contain all three types of information. For example, raster maps usually do not include attribute information, and many vector data sources do not include display information.

Geographic Information

The geographic information in a digital map provides the position and shape of each map feature. For example, a road map's geographic information is the location of each road on the map (Thomas 1970).

In a vector map, a feature's position is normally expressed as sets of X,Y pairs or X,Y,Z triples, using the coordinate system defined for the map (see the discussion of coordinate systems, below). Most vector geographic information systems support three fundamental geometric objects:

- Point: A single pair of coordinates.
- Line: Two or more points in a specific sequence.
- Polygon: An area enclosed by a line.

Some systems also support more complex entities, such as regions, circles, ellipses, arcs, and curves.

Attribute Information

Attribute data describes specific map features but is not inherently graphic. For example, an attribute associated with a road might be its name or the date it was last paved. Attributes are often stored in database files kept separately from the graphic portion of the map. Attributes pertain only to vector maps; they are seldom associated with raster images.

GIS software packages maintain internal links tying each graphical map entity to its attribute information. The nature of these links varies widely across systems. In some the link is implicit, and the user has no control over it. Other systems have explicit links that the user can modify. Links in these systems take the form of database keys. Each map feature has a key value stored with it; the key identifies the specific database record that contains the feature's attribute information.

Should problems arise, it is important to know how your software establishes and maintains attribute links (USA 1967)

Display Information

The display information in a digital-map data set describes how the map is to be displayed or plotted. Common display information includes feature colors, line widths and line types (solid, dashed, dotted, single, or double); how the names of roads and other features are shown on the map; and whether or not lakes, parks, or other area features are color coded.

However, many users do not consider the quality of display information when they evaluate a data set. Yet map display strongly affects the information you and your audience can obtain from the map - no matter how simple or complex the project. A technically flawless, but unattractive or hard-to-read map will not achieve the goal of conveying information easily to the user.

Oddly enough, many common data sets contain no display information. For example, USGS Digital Line Graph files provide no display information at all. Each feature contains an attribute that describes the entity but does not indicate display features. Users, and their GIS software, must interpret those attributes and decide how each will look on the final display (Neuner 2000).

3. CONCLUSIONS AND FUTURE WORK

The paper concentrates solely on providing fundamental information about specific most used data file formats in GIS applications. The aim of the paper is to propose an overview of the problem of geographic information knowledge representation and management, in order to familiarize and focus to the one of the most dynamic family of multimedia tools.

The next objectives of the authors are in the area of identification of the most adapted software sollutions for Romanian surveilling market. The aim of this approach is to promote regional information as a modern presence based on the best software sollutions and data representing and interpreting.

4. REFERENCES

Defense Mapping Agency. (1977). The American Practical Navigator Publication No. 9, Defense Mapping Agency Hydrographic Center.

Defense Mapping Agency. (1991). World Geodetic System 1984 (WGS 84) - Its Definition and Relationships.

Muehrcke, P.C and Juliana O. Muehrcke. (1992). Map Use. Madison, WI: JP Publications.

Neuner, J., (2000), Sisteme de pozitionare globala, MatrixRom, Bucuresti.

Snyder, John P. (1987). Map Projections, A Working Manual. Washington, DC: US Govt. Printing Office.

Thomas, P. D. (1970). Spheroidal Geodesics, Reference Systems and Local Geometry. Washington, DC: U. S. Naval Oceanographic Office.

U.S. Army. 1967. TM 5-241-1 Grids and Grid References. Washington, DC: Department of the Army.